



King's Research Portal

DOI:

[10.1109/TMM.2019.2936305](https://doi.org/10.1109/TMM.2019.2936305)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Liu, X., Dohler, M., & Deng, Y. (2020). Vibrotactile Quality Assessment: Hybrid Metric Design Based on SNR and SSIM. *IEEE TRANSACTIONS ON MULTIMEDIA*, 22(4), 921-933. [8807345].
<https://doi.org/10.1109/TMM.2019.2936305>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Vibrotactile Quality Assessment: Hybrid Metric Design Based on SNR and SSIM

Xun Liu, *Student Member, IEEE*, Mischa Dohler, *Fellow, IEEE*, and Yansha Deng, *Member, IEEE*

Abstract—The emerging mulsemmedia (MULTiple SENSORIAL MEDIA) introduces new sensorial data (haptic, olfaction, gustation, etc.), significantly augmenting the conventional audio-visual communication. This can be used in many areas, such as immersive entertainment and innovative education. Previous research has been dedicated to evaluating the impact of other sensorial data on conventional multimedia; however, standalone quality evaluation of new sensorial data, especially vibrotactile data (a type of haptic data), has not been covered. To the best of our knowledge, this paper is the first to empirically demonstrate that the common statistical metrics in audio and visual domains, i.e. signal-to-noise ratio (SNR) and Structural SIMilarity (SSIM), are highly correlated with human vibrotactile perception as well. To be specific, we propose a testing protocol for vibrotactile quality evaluation and conduct subjective experiments. The results suggest that SNR and SSIM are applicable to vibrotactile quality assessment. We also consider a practical scenario where the quality of vibrotactile data varies with time. Based on the validation of SNR and SSIM in the first part, we present an objective metric as a hybrid composition of SNR and SSIM. Instead of assessing the quality of data using an overall score, the hybrid metric evaluates the quality in a time-varying manner. Subjective experiments are conducted and the results demonstrate that the correlation coefficient can be significantly increased using the hybrid metric.

Index Terms—Mulsemmedia, haptics, vibrotactile, quality of experience, subjective tests, objective metrics.

I. INTRODUCTION

IN recent years, mulsemmedia has attracted enormous attention since it was firstly proposed in [1]. Apart from conventional audio-visual content, the mulsemmedia commits to transmit other sensorial data (haptic, olfaction, gustation, etc.), so that users can get truly immersive experience [2]. The introduction of the multiple sensorial data raises the issue of assessing the impact and quality of these data. The authors of [3] presented a subjective quality assessment method (based on ‘Absolute Category Rating with Hidden Reference’) to investigate how sensory effects (e.g. vibration, wind, light) influenced human perceived quality of video. The authors of [2] proposed a mulsemmedia framework to deliver both conventional multimedia and other sensorial data, and conducted subjective experiments to assess the perceived quality and enjoyment levels. Similar tests were conducted and human

perceived quality of 3D video accompanied by multisensorial components were evaluated in [4]. The research investigated the impact of multisensorial data (haptic, olfaction, gustation, etc.) on conventional multimedia in terms of human perception. They are based on the fact that quality evaluation methods of traditional multimedia are well established; therefore, it is essential to develop standalone quality assessment for the multisensorial data as well. The authors of [5] proposed a model for assessing quality of experience (QoE) of olfaction in the context of mulsemmedia, while the impact of scent type on QoE of olfaction was evaluated in [6]. Currently, there are few quality assessment methods dedicated to haptic data. During the processing, compression, transmission, reconstruction and synchronisation, haptic data experience various degradation at each stage; therefore, it is essential to quantitatively evaluate the quality of haptic data with respect to human perception.

Haptic information can be categorised into kinesthetic and tactile modalities. The former refers to perception of body movements and the latter stands for perception of mechanical or thermal stimuli sensed by the receptors in the skin. Kinesthetic signals normally have large amplitude and low frequency, while vibrotactile signals (a type of tactile signals used for texture discrimination) have small amplitude and high frequency. As for latency requirements, kinesthetic signals are strict [7], whereas vibrotactile signals are more tolerant [8]. In addition, the neural cells that are responsible for kinesthetic and vibrotactile perception are different, hence the perceptual properties and limits are distinct as well. Consequently, the evaluation of kinesthetic signals and vibrotactile signals should differ. Previous studies (e.g., [9]–[13]) have primarily concentrated on kinesthetic signals, while vibrotactile signals draw limited attention. This article focuses on the quality assessment of vibrotactile signals. To evaluate the quality of vibrotactile data, one can either conduct subjective tests or apply objective metrics.

Subjective evaluation, as the most reliable approach of quality assessment [14], normally requires a certain group of subjects to assess the quality under certain conditions. The accuracy of subjective evaluation heavily depends on the subjects’ performance; therefore, every step of subjective evaluation requires cautious design to ensure that the results reflect the subjects’ actual perception as close as possible. To date, only a little research has been dedicated to subjective evaluation of vibrotactile data. For instance, the authors of [15] conducted experiments to assess the quality of mulsemmedia; however, they only investigated the impact of time delay caused by synchronisation while the intrinsic quality of vibrotactile data was not considered. The authors of [16] and [17]

X. Liu, M. Dohler and Y. Deng are with the Department of Informatics, King’s College London, Bush House, 30 Aldwych, London, WC2B 4BG United Kingdom e-mail: (xun.2.liu@kcl.ac.uk; mischa.dohler@kcl.ac.uk; yansha.deng@kcl.ac.uk).

This paper has supplementary downloadable material available at https://drive.google.com/open?id=1ydrBOFJBESnR_arXWKDftbL4407Hwh1T, provided by the author. The material includes the texture data used in this paper. This material is 488 KB in size.

proposed a vibrotactile codec, respectively. For assessing the performance of the proposed codecs, they conducted subjective experiments using different methods; however, these methods are particularly designed for their codecs. In other words, these methods cannot be widely used for all vibrotactile applications.

In practice, objective metrics are more convenient, less time-consuming and less expensive than subjective evaluation. According to the availability of a reference signal, the objective metric is classified into full-reference, no-reference and reduced-reference. Although the latter two are more applicable in practical scenarios where reference signals are not always available, they normally confine to certain constraints [18]. As the design of vibrotactile objective metric is in the initial stage, this article focuses on full-reference objective metric that is fundamental. Currently, there is very limited literature with respect to vibrotactile objective metrics. The authors of [16] proposed an objective index that could estimate the quality of distorted signals with an accuracy of about 80% in [19]. This index, however, was particularly designed for the codec in [16], preventing it to be used in other applications. The authors of [10] designed an objective metric for haptic signals based on Weber's law; however, they considered only force, position and velocity feedback, leaving vibrotactile feedback out. In summary, there are no objective metrics dedicated to vibrotactile signals to date.

Instead of starting from scratch, it is possible to borrow insights from the well-developed objective metrics for other media. The authors of [17] presented that vibrotactile signals were similar to speech signals. They proposed that coarse textures could be treated as voiced sounds and fine textures could be treated as unvoiced sounds. The authors of [20] further demonstrated that vibrotactile and audio signals were similar from the perspective of mechanism and representations. Moreover, the authors of [21] empirically demonstrated that vibrotactile signals had similar masking phenomenon as audio signals and successfully adapted an audio codec to the vibrotactile domain; hence, there is potential to adapt audio quality assessment methods to vibrotactile quality evaluation as well. Motivated by the similarity between vibrotactile and audio signals, we design a hybrid objective metric based on SNR and SSIM, and conduct subjective experiments to assess the performance of it. The contributions of this article can be summarised in the following points:

- Through studying the standardised protocols in the audio domain, we adapt the most appropriate one, i.e. Rec. ITU-R BS.1534-3, to a subjective protocol that is exclusive for the vibrotactile quality evaluation.
- Based on this protocol, we design a subjective test to evaluate the correlation of SNR and SSIM with human vibrotactile perception. The results demonstrate that SNR and SSIM are highly correlated with human perceived quality.
- To conduct the subjective test, we design a platform that allows the participants to individually and conveniently assess the quality of vibrotactile data. The software is extensible and easily modifiable; therefore, it has potential to be used for other vibrotactile quality evaluation tests in the future.

- Considering a realistic scenario with distortion varying with time, we design a hybrid metric by combining SNR and SSIM in the spirit of Perceptual Evaluation of Audio Quality (PEAQ). Subjective tests are conducted and our results show that the proposed hybrid metric significantly outperforms SNR or SSIM.

The remainder of the paper is organised as follows. In Section II, we explicitly compare subjective methodologies and objective metrics in the audio domain and choose the ones that are possibly adapted to vibrotactile quality assessment. Besides, the design of the hybrid metric is presented in this section. In Section III, we demonstrate the collection, processing and display of vibrotactile data as well as the testing interface and procedure, while a subjective protocol for vibrotactile quality assessment is proposed in Section IV and experiments are conducted to investigate the correlation of SNR and SSIM with human vibrotactile perception. In Section V, we present the experiments for the time-varying scenarios, while the performance of the objective metrics are discussed in Section VI. Finally, we conclude the paper in Section VII.

II. VIBROTACTILE QUALITY ASSESSMENT

Over the past decades, quality evaluation for audio and video content has been well developed based on full understanding of auditory and visual psychophysics. Although many properties of touch await explorations, the mechanism of vibrotactile perception is clear. Consequently, we can learn from audio and image quality evaluation and adapt methods that are most appropriate for vibrotactile quality evaluation.

A. Vibrotactile Perception

Humans can feel objects and the environment because the neural cells in the skin – known as receptors in neuroscience – get excited by physical stimuli. The excitation is then converted into electrical signals followed by being transmitted to the brain for analysing and generating perception. This process is similar to that of other human sensory systems, such as hearing and vision, except that the type and number of receptors are different. Specifically, there are four types of receptors in the glabrous skin that are responsible for the sense of touch. They are called mechanoreceptors because they are sensitive to mechanical movements and deformation [22]. For instance, the Pacinian corpuscle responds to vibrations (up to 1000 Hz) and the most sensitive frequency is from 250 to 550 Hz [23]. The detailed characteristics and functions of these mechanoreceptors are listed in Table I. We can see that the mechanoreceptors get excited by different types of stimuli and have various ranges of sensitive frequency. Humans rely on a combination of a part or all of them to accomplish tasks that are as simple as holding an object properly or as complicated as writing in the darkness.

Interestingly, apart from directly touching objects, humans can actually perceive objects through intermediate tools because humans can feel the vibrations transmitted over the tools [24], [25]. For example, people use a key to open the door and chefs use a knife to properly handle ingredients. It

TABLE I
FUNCTION, ROLES AND SENSITIVE FREQUENCY OF FOUR TYPES OF
MECHANORECEPTORS (ADAPTED FROM [22]).

	Merkel cell	Ruffini ending	Meissner corpuscle	Pacinian corpuscle
Most sensitive	Pressure, edges, corner, points	Stretch	Lateral motion	Vibrations
Role	Reading Braille	Holding large objects	Sensing slippage of objects	Sensing texture
Frequency range (Hz)	0 – 100	/	1 – 300	5 – 1000

is, however, not easy to accurately and appropriately capture the vibrations. Conventional haptic systems always lack the realistic feeling of remote targets because the texture and other properties of the objects are only virtually rendered, without being measured in the real world. For enhancing the realism in virtual environments, the authors of [26] built a reality-based model of vibrations by recording the acceleration of a stylus tapping on real surface. The authors of [27] used a similar approach and demonstrated that high frequency signal can dramatically improve the contact realism of haptic systems. Furthermore, the users were allowed to record the acceleration signal by using a handheld tool to slide on the surface of objects [28]. In this study, experiments were conducted and results showed that high frequency vibrational signal is one of the main cues for humans to discriminate different materials. Therefore, it is the acceleration signal that needs to be evaluated for vibrotactile quality assessment.

B. Subjective Testing Protocol

Retrospecting the development of audio and visual quality evaluation, neither subjective nor objective methods can fully reflect true human perception; however, subjective evaluation can help us to get estimates as close as possible of actual perception. Although conducting subjective tests is time- and money-consuming, it is the most reliable way of assessing quality if it is carefully designed. On the contrary, objective metrics are just mathematical indicators and hence are convenient to be used in a variety of applications. Researchers endeavour to design objective metrics thoroughly so that objective metrics match human actual perception; however, this cannot be accomplished directly. The mapping can only be validated through subjective tests. The relation among human actual perception, subjective evaluation and objective metrics is illustrated in Figure 1.

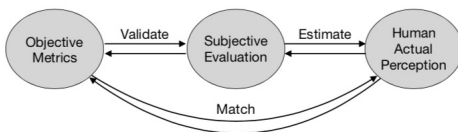


Fig. 1. Relation of actual human perception, subjective evaluation and objective metrics.

To the best of our knowledge, there is no standardised subjective testing protocol for assessing vibrotactile quality to date, but several methods exist for evaluation of audio quality. Three state-of-art subjective methods are standardised and recommended by the International Telecommunication Union (ITU) as follows: i) ITU-T P.800; ii) Rec. ITU-R BS.1116-3; iii) Rec. ITU-R BS.1534-3. The details of these protocols can be found in [29], [14] and [30], respectively. As mentioned before, vibrotactile and audio signals are similar in several aspects, which inspires us to modify one of the methods shown above.

ITU-T P.800 is for assessing telephone transmission quality with the well-known mean opinion score (MOS), which scales from 1 to 5 corresponding to "bad" to "excellent" as for quality. This approach is designed for evaluating quality of speech signals by simulating the typical scenario of a phone call; therefore, subjects are not able to directly compare the distorted signal with the pristine signal during the conversations. In other words, the reference signal is not always available for comparison. Instead, subjects can only compare the test signal with the undistorted signal in their memory. The reason behind this design is that humans are good at remembering the natural sound of human voice. Nonetheless, this is not the case for the vibrotactile quality assessment since the majority of the participants in our pilot tests report that it is very hard to remember vibrotactile signals.

Rec. ITU-R BS.1116-3 is standardised for the evaluation of small impairments in audio systems. Specifically, small impairments refer to distortions that can only be perceived by experts with statistical analysis under rigorous conditions. The target of ITU-R BS.1116-3 is extremely high quality audio systems. In contrast, the design of vibrotactile quality assessment is in the initial stage. The primary goal should be assessing a large range of vibrotactile impairments rather than focusing on the hard-to-detect areas. It is not only a waste of time and efforts but also possible to yield unreliable reports if ITU-R BS.1116-3 is utilised for assessing obvious impairments. In addition, another practical reason of not using this approach is that it strictly requires "subjects who have expertise in detecting small impairments", but it is almost impossible to find enough expert vibrotactile assessors for now.

On the contrary, the intention of Rec. ITU-R BS.1534-3 is to assess intermediate quality in audio systems. This approach employs a strategy called "MULTi Stimulus test with Hidden Reference and Anchor (MUSHRA)". From the name, it is apparent that there are hidden reference signals and anchors apart from the test signals. The anchors are impaired signals that are carefully designed for certain situations. For instance, the standard anchors in the protocol are generated by applying low-pass filters to the pristine signal with a cut-off frequency of 3.5 kHz and 7 kHz, respectively. To avoid speculation from subjects, test signals along with a hidden reference signal and anchors are randomly deployed. The subjects are allowed to listen to any of the signals as many times as they want in a single trial before scoring each signal. The scores are given on the basis of the continuous quality scale (CQS), which ranges from zero to one hundred and is equally divided into 5 intervals

referring to "Bad", "Poor", "Fair", "Good" and "Excellent" in terms of quality.

TABLE II
FEATURES OF THE STANDARDISED SUBJECTIVE TESTING PROTOCOLS FOR AUDIO SIGNALS

Protocol	Target	Parallel comparison	Grading system	Selection of subjects
ITU-T P.800	Telephone transmission	Not available	Mean opinion score (1-5)	Normal telephone users
ITU-R BS.1116-3	Small impairments	Not available	Continuous five-grade impairment scale (1-5)	Expert
ITU-R BS.1534-3	Intermediate impairments	Available	Continuous quality scale (0-100)	Experienced

The features of the standardised audio protocols are summarised in Table II. Three pilot tests were conducted with three methods used in the presented protocols. The Comparison Category Rating (CCR) method in ITU-T P.800 and the double-blind triple-stimulus with hidden reference method in ITU-R BS.1116-3 were not chosen because they required expert assessors, which were beyond the scope of vibrotactile quality assessment at this stage. Three methods used in the pilot tests are:

- The Absolute Category Rating (ACR) method: The subjects perceived the reference signal and then a group of distorted signals in a random order. Each signal was perceived only once and the subjects were asked to give scores right after perceiving each stimulus. The subjects reported that it was hard to give scores when the tests came to the third or fourth stimulus. They mentioned that the memory of reference signals faded quickly and previous test signals further disturbed the memory of reference signals.
- The Degradation Category Rating (DCR) method: The subjects were presented with two signals in pairs, with the first one always being the reference signal and the second one being an impaired signal randomly chosen from the group of distorted signals. A short interval was inserted between two signals so that the subjects could distinguish two stimuli. Each pair was perceived as many times as the subjects preferred. The subjects gave scores to the second signal after perceiving each pair of signals. The subjects suggested that it would be more accurate to grade distorted signals that stimulated close perception if parallel comparison between distorted signals was allowed.
- The multi stimulus test with hidden reference method: This one was a preliminary version of the method used in this paper. There were seven test signals in one trial and each test signal lasted for two seconds. No subjects reported that it was hard to give scores to the stimuli; however, some subjects reported numbness of the fingers at the later stage of the test. Hence, we carefully reduced

the duration of the signals and the number of test signals for the formal tests.

It turns out that Rec. ITU-R BS.1534-3 suits our current needs for assessing vibrotactile quality because it has the following advantages:

- The pristine signal is always accessible so that subjects can compare test signals with the reference signal as long as they want in a single trial, which is significant since the pilot tests suggest that humans are weak in recalling vibrotactile information.
- Parallel comparison is allowed in this protocol. To be specific, subjects are able to compare two distorted signals directly, thus discriminating and scoring them easily and precisely.
- Rec. ITU-R BS.1534-3 considers impairments that are detectable to experienced users instead of expert assessors which are not available at present. Besides, assessing intermediate impairments accords with the current goal of vibrotactile quality assessment and facilitates the validation of vibrotactile objective metrics.
- Compared with the widely used five-score MOS, the CQS that has an extent from zero to one hundred is capable of reflecting the inevitable variances between subjects. Consequently, the results obtained in this way are more convincing.

However, it is not possible to directly employ Rec. ITU-R BS.1534-3 as it is designed for audio quality evaluation. In order to implement this method in the vibrotactile domain, we made several alterations with respect to human vibrotactile perception, the details of which will be shown in Section IV.

C. Objective Metrics

The designs of full-reference objective metrics in audio and visual domains are classified into three categories. The first one is to consider the impact of noise in a pure statistic way. To be specific, the power or other properties of the noise are mathematically calculated and compared with that of the pristine signal. The most commonly used ones are SNR in the audio domain and peak signal-to-noise ratio (PSNR) in the visual domain. Besides, [31] presented another statistical method to evaluate the quality by measuring the SSIM between pristine and distorted images. The second approach is based on the psychophysical model of human perception. Researchers either incorporate perceptual models with statistical measures, e.g., Visual Signal-to-Noise Ratio (VSNR) [32] in the visual domain, or extract multiple perceptual features to build an index, e.g., Information Fidelity Criterion (IFC) [33] in the visual domain and Enhanced Modified Bark Spectral Distortion (EMBSD) [34] in the audio domain. The last approach is hybrid metrics that coordinate all the useful information of pristine and distorted signals as well as complete psychophysical models to build an overall grade. The state-of-art hybrid metrics in the audio domain are Perceptual Evaluation of Speech Quality (PESQ) and Perceptual Evaluation of Audio Quality (PEAQ), standardised in ITU-T P.862 [35] and Rec. ITU-R BS.1387-1 [36], respectively.

The approaches described above also illustrate the development history of the objective metrics in the audio and visual domains. The researchers firstly design the statistical approaches that can be used for certain applications. As more psychophysical properties were discovered, more complicated codecs have been designed according to human perceptual models. The pure statistical metrics are no longer able to accurately assess the performance of such codecs; hence, the latter two types of metrics are proposed for complicated tasks. All of these are based on full understanding of human auditory and visual systems; however, the psychophysics of touch are not completely understood. Furthermore, the complexity of the objective metrics dramatically rises as the number of perceptual models included increases. On the contrary, the statistical approaches have advantages in simplicity and expandability as some of the psychophysical approaches actually rely on the statistical metrics. Consequently, it would be reliable and practical to follow the path of the development of objective metrics in audio and visual domains. It is not only necessary but also critical to validate the statistical metrics in the vibrotactile domain. In this article, the widely-used SNR and SSIM are chosen.

1) *SNR*: SNR is defined as the power ratio of pristine signal to noise. As the dynamic range of the signal is wide, the SNR is normally expressed using the logarithmic decibel scale. The SNR in decibel is calculated as follows:

$$\text{SNR} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right), \quad (1)$$

where P_{signal} and P_{noise} are the power of the pristine signal and noise, respectively.

Similarly, the PSNR is the power ratio of the maximum possible value to noise. While SNR is widely used in the audio domain, PSNR is commonly used in the visual domain as a pixel is normally represented using a fixed number of bits, e.g., 8 bits for grayscale images and 24 bits for true colour images. Apparently, this is not the case for vibrotactile signals that are similar to audio signals, hence we decided to validate the SNR rather than PSNR for vibrotactile quality assessment. From Eq. 1, we can see that the larger the SNR is, the smaller the power of noise is for one pristine signal; however, this does not necessarily mean that human perceived stimulus is weaker. Even if the SNR is demonstrated to be in accordance with the human sensation of touch, we need to figure out the strength of the correlation. Only in that case, we are able to see if SNR is applicable to vibrotactile quality assessment.

2) *SSIM*: On the other hand, SSIM takes into account the structural relation between pristine and distorted signals rather than the power of noise. Utilising the mean, standard deviation and covariance of both signals, SSIM computes three factors as:

$$L = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2)$$

$$C = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (3)$$

$$S = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (4)$$

where x and y are pristine and distorted signals, respectively. μ , σ and σ_{xy} are the mean, standard deviation and covariance, respectively. C_1 , C_2 and C_3 are small constants to prevent the equation from instability when the denominator is zero. L , C and S are defined as luminance comparison, contrast comparison and structural comparison, respectively. Similarly, we denote these factors as roughness comparison, stiffness comparison and structural comparison in the vibrotactile domain. These factors are then combined to build a single index that measures the similarity between pristine and distorted signals [31]. It is defined as:

$$\text{SSIM} = L \cdot C \cdot S. \quad (5)$$

As each factor is from -1 to 1 , the integrated SSIM is within the interval $[-1, 1]$. SSIM holds the property of identity, which means the absolute value of SSIM equals to 1 if and only if x and y are identical. In other words, no noise is introduced. In practice, it is more precise to calculate the SSIM locally using a sliding window, then compute the average of the SSIMs to get an overall value, named mean SSIM (MSSIM), which is expressed as:

$$\text{MSSIM} = \frac{1}{N} \sum_{i=1}^N \text{SSIM}_i, \quad (6)$$

where N is the number of windows and SSIM_i is the SSIM of the i -th window. The SSIM is originally designed for image quality evaluation [37]. It is, however, a pure mathematical metric without taking any psychophysics of the human visual system into account, implying the potential to apply it to other human senses. Recent studies justified this hypothesis. The authors of [38] proposed two variants (Temporal-MSSIM and Time-Frequency-MSSIM) of SSIM and demonstrated that the variants had high correlation with human auditory system. Temporal-MSSIM and Time-Frequency-MSSIM, despite of being variants of SSIM, remained the characteristic of being pure mathematical. Besides, SSIM captures structural information (dependencies among neighboured samples), which is demonstrated to be important in the visual domain. The structural information is also vital in the vibrotactile context, since the change of structure that results in small change of SNR may lead to large differences in terms of human vibrotactile perception. Hence, we exploit the potential of applying SSIM for vibrotactile quality assessment in this paper. In order to check the feasibility of SSIM as well as SNR, we propose the first test and present it in Section IV.

Despite the fact that we need to demonstrate that SNR and SSIM are suitable to vibrotactile quality assessment, we can take a step further and figure out which one is better with respect to the human vibrotactile perception. In terms of complexity, SNR is lower than SSIM, but SNR does not hold the property of identity since it converges to infinity when the power of noise is extremely small. Besides, SNR is established on mean square error (MSE) that is widely used for solving optimisation problems. MSE acquires only the power of error signals (referring to noise signals in signal processing) which can be preserved during a Fourier Transform according to Parseval's Theorem. However, due

to the same reason, MSE cannot precisely reveal the impact of noise on pristine signals in terms of human perception. A simple example is that values of MSE obtained are the same when a positive and a negative noise signal with the same amplitude, respectively, is added to the same pristine signal; however, these distorted signals possibly lead to very different human perception. In contrast, SSIM is based on the fact that samples of signals are highly correlated with neighbour samples, which is demonstrated to be important for human perception [37]. From the perspective of statistics, SSIM contains more information than SNR as well. The authors of [39] showed that SSIM outperforms PSNR for image quality evaluation through quantitative comparison. The results of Test 1 also demonstrate that this situation holds for vibrotactile quality assessment.

3) *Hybrid metric design*: In practice, an objective metric ought to cope with more realistic situations because one may need to evaluate long sequences of vibrotactile data. Conventional SNR and SSIM intend to measure the quality of a sequence as a whole. However, it is inaccurate to give an overall score since the type and intensity of noise encountered possibly vary with time due to the instability of communication channel. Consequently, it would be more precise to piecewise evaluate the vibrotactile quality for a long time vibrotactile signal. Conventional SNR focuses on power measurement while SSIM is interested in structural information. Inspired by PEAQ, we designed a time-varying objective metric by taking advantages of both SNR and SSIM.

Currently, PEAQ is one of the most popular hybrid objective metrics in the audio domain, which combines several features, named model output variables (MOVs), extracted from the human auditory system to generate a single grade. In the light of PEAQ, SNR and SSIM are considered as MOVs in this article. The compounding process is illustrated in Fig. 2.

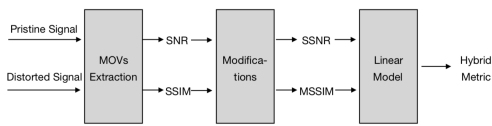


Fig. 2. Stages of creating the hybrid metric: SNR and SSIM are extracted from pristine and distorted signals and are modified as MOVs for compounding the hybrid metric.

As conventional SNR and SSIM produce only an overall grade for the whole sequence of data, the long vibrotactile signal is evenly divided into intervals and the SNR and SSIM of each interval are calculated to get a time-varying version. To be specific, a variant of SNR, named Segmental Signal-to-Noise Ratio (SSNR) [40] is used. SSNR segmentally calculates the SNR and then computes the average over all segments. In our case, the segment is chosen to be equal to the interval and the averaging step is left out. As to SSIM, we use MSSIM to compute the value for each interval and remove the averaging step. Consequently, a value of SSNR and MSSIM are obtained, respectively, for each interval of the long vibrotactile signal. According to Webber's law [41], human sensation is not linear,

hence we apply a 3rd order polynomial curve fitting as follows:

$$\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 = y_i \quad (i = 1, 2, 3, 4), \quad (7)$$

where x_i is the value of SSNR or MSSIM in the i th interval and y_i is the subjective score in the i th interval. By solving this linear equation, we can get the optimal coefficients $(\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*)$. We then obtain the modified SSNR and MSSIM as:

$$z_i = \beta_0^* + \beta_1^* x_i + \beta_2^* x_i^2 + \beta_3^* x_i^3 \quad (i = 1, 2, 3, 4). \quad (8)$$

Then, we build a matrix $A \in \mathbb{R}^{M \times N}$ of time-varying metrics, where M is the Number of intervals and N is the number of MOVs. The element, $A_{i,j}$, in the matrix stands for the j th MOV for the i th interval. By setting the subjective scores of each interval as a column vector $B \in \mathbb{R}^{M \times 1}$, a linear equation is built as follows:

$$AW = B, \quad (9)$$

where $W \in \mathbb{R}^{N \times 1}$ is a vector that stands for the weights of each MOV. For this equation, we combine all of the subjective scores together to get a global $W_{Global} \in \mathbb{R}^{N \times 1}$ that is optimal for the whole testing dataset. Once the optimal W_{Global} is obtained using least-squares approximation, the hybrid metric is calculated as:

$$H = AW_{Global}. \quad (10)$$

III. EXPERIMENTAL SETUP

In order to implement the subjective tests, three problems must be addressed: i) how to properly collect, process and display vibrotactile data; ii) the software for conducting the experiments; iii) the testing protocol. We address the first two problems one by one in this section and the last one in Sections IV and V. The setup used is similar to the reference setup of the IEEE.1918.1.1 Standardisation Group [42]. The setup and flow diagram of the tests are illustrated in Fig. 3 and Fig. 4, respectively.

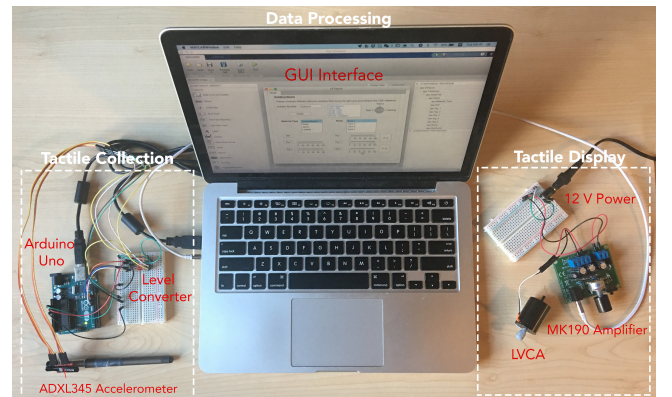


Fig. 3. Setup of the tests including vibrotactile data collection on the left, data processing and GUI Interface in the center and vibrotactile display on the right.

A. Vibrotactile Data Collection

In order to accurately capture the acceleration signals, we choose a digital accelerometer, referred to as ADXL345, from Sparkfun. It is a 3-axis accelerometer with a high resolution and a large range of measurements. In full resolution mode, the resolution can be as high as 3.9 mg/LSB and the measuring limit is up to $\pm 16g$. As the sensitive frequency of Pacinian corpuscle is up to 1000 Hz shown in Table I, the required Nyquist rate of vibrotactile signals is 2000 Hz according to the Nyquist-Shannon Sampling Theorem. The output data rate of the ADXL345 can be up to 3200 Hz which meets the requirement. To transmit the output data of the ADXL345 to a computer, we use an Arduino Uno that supplies 5V. As the ADXL345 supports only 2.0-3.6V, we add a logic level converter from Sparkfun for protection. As for the protocol of serial communications, the ADXL345 utilises either Serial Peripheral Interface (SPI) or Inter-Integrated Bus. The 4-wire SPI mode is chosen as it does not require any external components, thus having high capability of resisting disturbance.

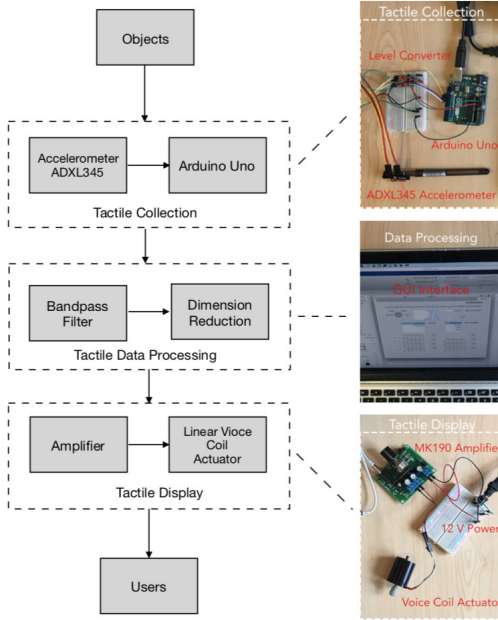
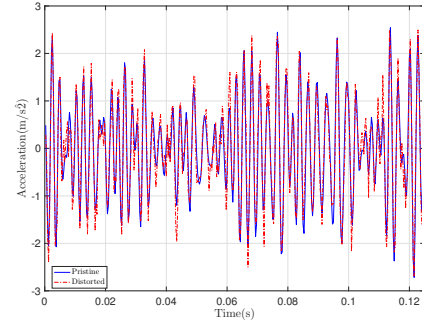


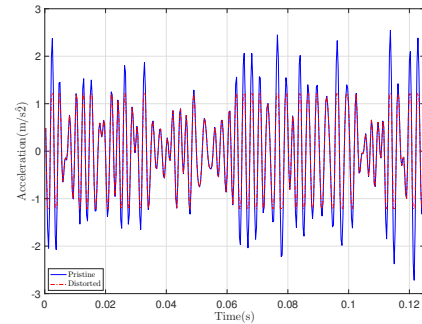
Fig. 4. Flow diagram of the tests: the vibrotactile data is collected using an accelerometer, followed by processing, then are displayed to users through a linear voice coil actuator.

As mentioned in Section II, humans heavily rely on perceiving texture of objects through tools; therefore, we record the acceleration signals in this way. Firstly, we firmly attach the ADXL345 to a pen. For collecting the vibrotactile data, we scan the surface of objects by freely moving the pen back and forth between two points. The purpose of this is to imitate human natural movements of exploring objects. A typical waveform of the acceleration signal is shown in Fig. 5. The duration of collected acceleration signals is one second and four seconds for the tests in Sections IV and V, respectively. As for selection of materials, it would be ideal to use as many materials as possible; however, that would dramatically increase the duration of the whole test. As the

duration rises, the numbness felt by the subjects intensifies, which is likely to cause inaccurate results. For this practical concern, we carefully choose four types of materials that cover "soft and coarse", "soft and fine", "hard and coarse" and "hard and fine", to reflect the variations of different materials (the texture data can be found in supplementary materials). The details and pictures of the chosen materials are illustrated in Fig. 6.



(a) Adding GWN with SNR = 10



(b) Adding LUQN with SNR = 10

Fig. 5. Pristine and distorted vibrotactile signals of *Coarse Board* with adding GWN and LUQN, respectively, where GWN = Gaussian white noise and LUQN = linear uniform quantisation noise.

B. Data Processing

After collecting the vibrotactile data of four types of materials, a Bessel bandpass filter is used, as the Pacinian corpuscles are most sensitive to vibrations with frequency ranging from 250 to 550 Hz [23]. Besides, the acceleration signals obtained are three dimensional while the vibrotactile display can only present one-axis vibrations; hence we apply a dimension reduction algorithm proposed by [43]. With this algorithm, the vibrotactile data can preserve temporal and spectral properties when it is transformed from three dimensions to one. In other words, no perceivable degradation is introduced after the transformation. After these processing, we combine the one-dimensional filtered vibrotactile data for all the materials together to build a pristine vibrotactile dataset.

However, for evaluating performance of SNR, SSIM and the hybrid metric, it is insufficient to only have the pristine vibrotactile dataset as the subjects need to give scores by perceiving both pristine signals and distorted signals. Hence, we build a distorted vibrotactile dataset for each test by adding

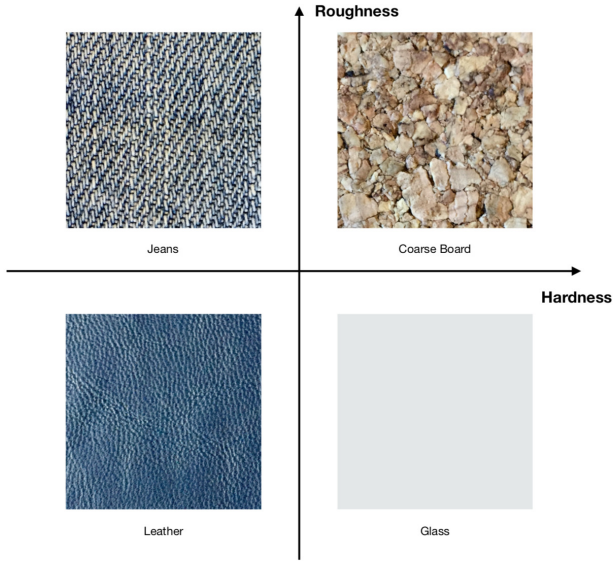


Fig. 6. Material selection and respective properties: four materials that cover "soft and coarse", "soft and fine", "hard and coarse" and "hard and fine" are chosen to show variances of the material.

different types of noise with various intensities (i.e. various SNR and SSIM) to the pristine vibrotactile data. Since the testing protocols of the tests in Sections IV and V are different, the type of noise added are different. The generation of the distorted vibrotactile dataset used in each test will be described in Sections IV and V, respectively.

C. Vibrotactile Display

For delivering vibrational signals, linear actuators rather than the widely used rotary actuators or motors are considered. Currently, there are many types of linear actuators, such as mechanical actuators, hydraulic actuators, pneumatic actuators and linear motors; however, none of them are capable of providing high frequency vibrations. By contrast, linear voice coil actuators (LVCA) driven by the electromagnetic force can not only produce high frequency vibrations, but are also highly controllable. According to Lorentz force law, the force of LVCA is calculated as follows:

$$F = k \cdot B \cdot L \cdot I \cdot N \quad (11)$$

where k is the force constant, B is the magnetic flux density, L is the length of coil, N is the number of coils and I is the current passing through the coil. For a given LVCA, the previous four parameters are fixed, which means the force is proportional to the current. The acceleration is $a = F/m$, where m is the mass of the moving part. In another word, we can control the acceleration by regulating the current passing through the coil.

Model NCM02-10-008-2JBA from H2W technologies [44] is chosen for these tests. The peak force of it is 11.2 N and the moving mass is 0.024 kg; the maximum acceleration is thus approximately 467 m/s^2 . This limit is enough for our purpose since the maximum acceleration of coarse material

is only about 3.5 m/s^2 . In order to drive the actuator, we employ an amplifier MK190 from Velleman [45]. Besides, as the small shaft of LVCA is hard to hold, the subjects possibly miss some vibrations, which potentially leads to unreliable results. We create a button-like holder using smart adhesives from Bostik [46]. This type of adhesives is similar to plasticine and can be firmly attached to the shaft. On the hand, it is harder than plasticine, thus providing solid touch. In the tests, the subjects are asked to hold the button with three fingers for firm and full contact.

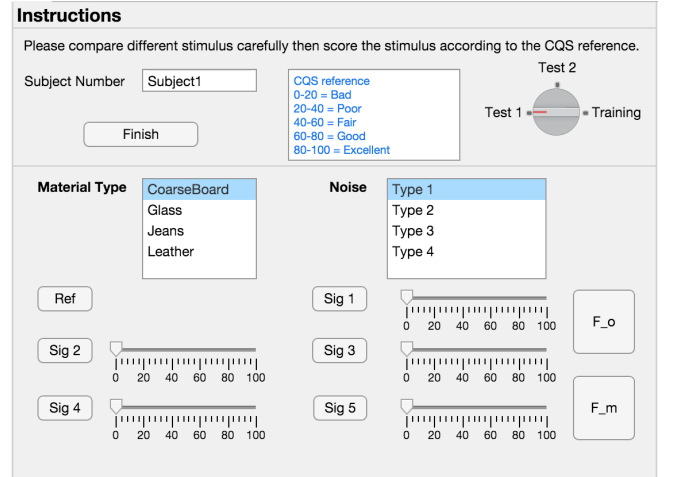


Fig. 7. GUI Interface of MUSHRT: Users can choose the number of the test, type of material and noise using the knob on the right-top corner and the list in the centre, respectively. The signal button is for playing vibrotactile data and the slider nearby is for scoring. F_o and F_m are for playing long sequence pristine and distorted signals in Test 2, respectively.

D. GUI Interface

Currently, there are some software for MUSHRA in the audio domain, but it is not suitable for vibrotactile quality assessment; therefore, we design a GUI Interface (see Fig. 7) using Matlab according to the protocols presented in Sections IV and V. As the vibrotactile data are stored and processed in Matlab as well, there is no issue of compatibility. During the tests, the subjects can easily switch tests using the knob in the top right corner (Test 1 and Test 2 refer to the test in Sections IV and V, respectively). They can also choose a combination of one type of material and one type of noise using the item list in the centre. The score for each distorted signal is marked using the slider that is able to show variances of different subjects. With this GUI Interface, a subject, even with no experience of subjective tests can independently accomplish the tests, so that the impact of disturbance is reduced to a minimum. Furthermore, it is convenient to modify the software if one wants to alter the number and types of material or noise.

E. Testing Procedure

The testing setup is shown in Fig. 3. There are smart adhesives at the bottom of the LVCA, so that the subjects can firmly stick it wherever they feel comfortable prior to

the tests. As we need the subjects to give scores depending only on touch, the subjects are asked to wear a headphone to exclude the noise generated by the LVCA. At the beginning of the tests, the subjects are given a training session, in which the software and full range of the distorted signals are introduced. Thereby, they can get familiar with the software and have rough impressions of different levels of degradation. Then, they are allowed to freely explore any signals of the distorted dataset. They do not start the tests until they feel confident to give absolute scores of vibrotactile quality. For test 1 (Section IV), one type of material and one type of noise are chosen in one trial. The subjects then use six buttons ('Ref', 'Sig 1', 'Sig 2', 'Sig 3', 'Sig 4' and 'Sig 5') in the GUI interface (Fig. 7) to perceive different stimuli. Each button contains a vibrotactile signal that lasts one second (according to the subjective protocol presented in Section IV), with 'Ref' containing the reference signal and others containing the test signals (including a hidden reference signal). The test signals with different distortion are randomly mapped to the buttons of 'Sig 1' to 'Sig 5', so that a button with larger number does not necessarily contain a signal with larger distortion. This is to avoid speculation from the subjects. The reason for choosing five test signals is that, subjects of pilot tests (seven test signals) report strong numbness at later stage of the test. As the numbness may cause inaccurate results, we carefully reduce the number of test signals and find out that the resolution of five test signals is enough for assessing correlation. Next, they freely explore these buttons as many times as they want and use the slider behind the button to grade accordingly. After storing the scores, they move to the next type of material or next type of noise until they complete all combinations. For test 2 (Section V), one type of material and one type of noise pattern are chosen in one trial. The subjects use two buttons ('F_o' and 'F_m') to perceive long vibrotactile signals (four seconds according to the time-varying protocol presented in Section V), with the former button refers to reference signals and the latter refers to distorted signals. Then, the subjects give scores to each interval (one second according to the time-varying protocol presented in Section V). After finishing a trial, the subjects carry on until all combinations are conducted.

IV. TESTING DESIGNS FOR SNR AND SSIM

The objective of this test is to investigate the correlation between statistical metrics, i.e. SNR and SSIM, and human vibrotactile perception. The result of a subjective test is only valid if the test follows a proper protocol. According to the comparison in Section II, we adapted Rec. ITU-R BS.1534-3 to the vibrotactile domain in this section. With this protocol, we generated a distorted vibrotactile dataset and conducted subjective experiments to investigate the relation between statistical metrics and human vibrotactile perception.

A. Testing Protocol

As mentioned in Section II, Rec. ITU-R BS.1534-3 can not be directly applied to the vibrotactile domain. To design a

protocol that is dedicated to vibrotactile quality assessment, we made three alterations according to vibrotactile psychophysics:

- The duration of the test signals is changed. The recommended duration of the audio signal is quite long, i.e., ten seconds. The test materials used in the audio quality evaluation are normally music, vocal, speech, orchestra or artificial sound [47], which are friendly to humans. Thereby the experience of long-time listening is acceptable. By contrast, the vibrotactile signals presented to the subjects' fingers are essentially vibrations that not only lead to uncomfortable numbness but also generate unpleasant auditory noise. Such prolonged tests would increase the subjects' anxiety and hence largely reduce the accuracy of quality evaluation. As a result, the duration needs to be accordingly cut down. Through pilot tests, we found out that one second was sufficient for discriminating the intermediate impairments as well as keeping the feeling of numbness at an acceptable level, and that was why we collected one second of acceleration signals for each material in Section III.
- The protocol requires experienced assessors who "have experience in listening to sound in a critical way", but there are very few people who have experience in detecting impairments introduced to the vibrotactile signals and critically assessing the quality of the distorted vibrotactile signals using the vibrotactile display. The selection of assessors is critical for subjective tests as the results are convincing only if the subjects' grades are reliable. In order to get valid results, we make the following improvements: i) The training session is extended. After introducing the full range of impairments to the subject, the subjects are allowed to freely explore any of the stimuli until they feel confident of scoring. ii) The protocol requires less than twenty assessors, but we increased the sample size to thirty. iii) Significant testing is conducted to make sure the results are within the required resolution.
- We do not use any anchors in vibrotactile subjective tests. As mentioned, the anchors are carefully designed for representing certain quality. The main objective of using anchors is to compare the results of tests under different conditions. The anchors are only applicable if they are properly designed, otherwise they only introduce bias to the results. Actually, recent study shows that even the recommended anchors obtain different quality scores in different experiments, thus having the potential to introduce bias in MUSHRA Listening Tests [48]. In consequence, it is not appropriate to implement anchors for vibrotactile quality assessment at this stage.

Except for the alterations described above, other parts of Rec. ITU-R BS.1534-3 are adopted and details can be found in [30]. For convenience, we denote this protocol as "Multi Stimulus test with Hidden Reference for vibroTactile quality evaluation (MUSHRT)".

B. Distorted Vibrotactile Dataset for SNR and SSIM

The generation of distorted signals is shown in Fig. 8. We chose four types of common noise encountered in networks

and codecs, i.e. i) linear uniform quantisation noise (LUQN); ii) Gaussian white noise (GWN); iii) impulsive noise (IN); and iv) pink noise (PKN). LUQN is the common noise encountered during lossy compression. We use 8 bits uniform quantisation and change the step size to get a group of noise signals with SNR equalling to 1, 4, 7 and 10 (four SSIM values are obtained at the same time). GWN and IN are the most common noise used in information theory, while PKN is ubiquitous in biological systems and electronic devices [49]. For these types of noise, we change the variance to get groups of noise signals with different SNR values (SNR equals to 1, 4, 7 and 10). We define a combination of one type of noise and one type of material as one trial, therefore there are five test signals (including a hidden reference signal) in one trial. As there are four types of materials and four types of noise, a subject needs to score 80 signals through 16 trials in Test 1.

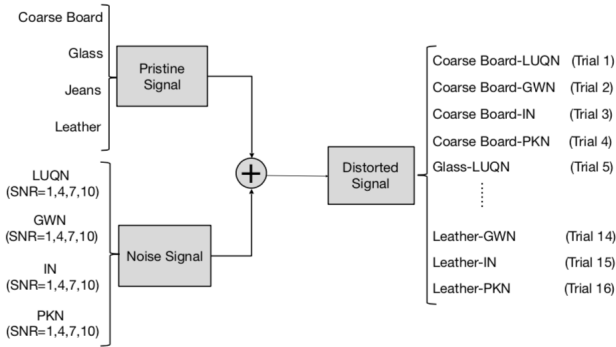


Fig. 8. Generation of distorted signals in Test 1: There are four noise signals that have different SNRs for one type of noise. There are four types of noise, i.e. LUQN = linear uniform quantisation noise, GWN = Gaussian white noise, IN = impulsive noise and PKN = pink noise. A combination of one type of material and one type of noise is defined as a trial, therefore a subject needs to take 16 trials in Test 1.

V. TESTING DESIGNS FOR THE HYBRID METRIC

A. Time-Varying Testing Protocol

As we need to evaluate the time-varying quality for a prolonged vibrotactile sequence, MUSHRT is no longer feasible. We design a new testing framework to imitate practical situations. The pristine signal is equally divided into intervals of one second and different types of noise with different intensity are then randomly added to each interval to create the distorted signal. During the test, the subjects need to score each interval while the entire sequence of pristine or distorted signal is displayed to the subjects at a time. In order to make sure the subjects exactly score each interval, a very short break (50 ms) is inserted between the intervals. Otherwise, the subjects possibly get confused about the timing to score if adjacent intervals present similar perception. The subjects are allowed to perceive the distorted signal as many times as they want while they are permitted to perceive the pristine signal only once. In other words, the subjects need to compare the distorted signal with the pristine signal in their mind, which is close to realistic scenarios.

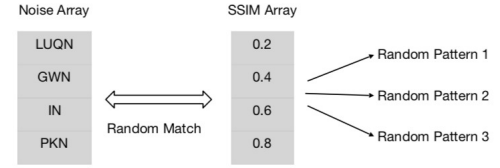


Fig. 9. Generation of random pattern: Each element of the noise array is randomly matched up with a value in SSIM array to create a random pattern, so each random pattern has four noise signals. We employ three random patterns for Test 2.

B. Distorted Vibrotactile Dataset for the Hybrid Metric

A four seconds pristine vibrotactile signal is employed and is evenly divided into four intervals. For generating random pattern of noise signals, we make a SSIM array ([0.2,0.4,0.6,0.8]) and a noise array (['LUQN','GWN','IN','PKN']). Each element of the noise array is then randomly matched up with a SSIM value to create a random pattern (see Fig. 9), hence a random pattern has four noise signals. Each noise signal is then randomly added to each interval of the pristine signal one by one to create the distorted signal. In this test, we use three random patterns, hence the subjects need to take 12 trials and give 48 scores in total. The whole process is shown in Fig. 10.

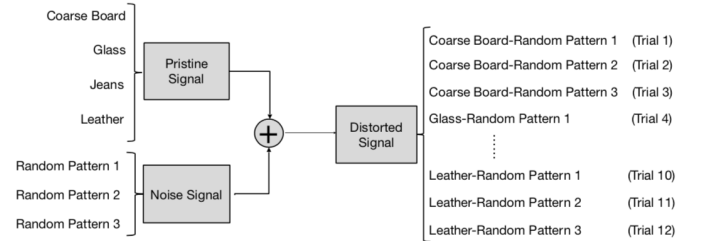


Fig. 10. Generation of distorted signals in Test 2: The process is similar to Test 1 except that the types of noise signals are changed to random patterns.

VI. RESULTS

Thirty subjects (17 female and 13 male) participated in the tests and all of them were right handed. They were all volunteers recruited through emails and none of them had vibrotactile impairments. Their ages ranged from 22 to 30, with an average of 26 years old. Only five subjects had taken part in vibrotactile psychophysical experiments before. The duration of tests ranges from 30 to 50 minutes. According to MUSHRT, we get the median of all the subjects' scores, then calculate the Pearson correlation coefficients between subjective scores and objective metrics. Using the correlation coefficient is a common way to evaluate the performance of an objective metric [12], [38]. The value of the correlation coefficient ranges from 0 to 1. The larger the value is, the stronger the relation between the two variables is.

A. SNR and SSIM

The correlation coefficients of four types of materials and four types of noise are shown in Table III and Table IV,

respectively. First, it can be seen from Table III that the correlation coefficients among different materials are very similar. The difference in correlation coefficients is less than 0.03 for either SNR or SSIM. As to roughness, the correlation coefficients of coarse materials (*Coarse Board* and *Jeans*) are slightly higher (around 0.01 to 0.02) than that of fine materials (*Glass* and *Leather*) for both SNR and SSIM. This suggests that both SNR and SSIM perform slightly better for assessing coarse materials. As to stiffness, SNR is more effective in evaluating hard materials (*Coarse Board* and *Glass*) than soft materials (*Leather* and *Jeans*), while it is on the contrary for SSIM. In Table IV, as for SNR, the correlation coefficient of PKN is much higher than that of other types of noise, while the correlation coefficient of LUQN is lower (approximately 0.08) for SSIM. It indicates that the SNR is slightly better at assessing signals affected by PKN compared with other types of noise, whereas SSIM shows degradation of evaluating signals affected by LUQN. Overall, the average correlation coefficients show that both SNR and SSIM are highly correlated with human vibrotactile perception (0.8391 and 0.9643, respectively), but SSIM has higher correlation coefficients (about 0.125 on average) than SNR. In other words, SSIM outperforms SNR in terms of assessing the quality of vibrotactile data.

For significance test, we choose the common two-tailed test. The degrees of freedom is computed as $N_d = N_s - 2$, where N_s is the number of samples. As there are 5 test signals in a trial, we get $N_s = 5$ and $N_d = 3$. According to the table of critical values for Pearson correlation coefficient, the statistical correlation coefficient is 0.8054 and 0.8783 at significance level of 0.1 and 0.05, respectively. Comparing with the average correlation coefficients in Table III (0.8391 and 0.9643 for SNR and SSIM, respectively), we can see that the correlation coefficient of SNR is significant at significance level of 0.1, while that of SSIM is significant at significance level of 0.05. In summary, it suggests that SSIM outperforms SNR for static vibrotactile quality assessment, which coincides with the observations in the audio and visual domains.

TABLE III
AVERAGE PEARSON CORRELATION COEFFICIENTS OF FOUR TYPES OF MATERIALS.

	SNR	SSIM
Coarse Board	0.8584	0.9656
Glass	0.8320	0.9566
Jeans	0.8459	0.9722
Leather	0.8201	0.9627
Average	0.8391	0.9643

TABLE IV
AVERAGE PEARSON CORRELATION COEFFICIENTS OF FOUR TYPES OF NOISE.

	SNR	SSIM
LUQN	0.8337	0.9066
GWN	0.8014	0.9742
IN	0.8223	0.9851
PKN	0.8990	0.9910
Average	0.8391	0.9643

TABLE V
THE OPTIMAL COEFFICIENTS OF CURVE FITTING.

	β_0^*	β_1^*	β_2^*	β_3^*
SSNR	21.3187	-1.0749	1.0285	-0.0461
SSIM	14.1329	77.0810	-195.5201	246.8429

TABLE VI
THE OPTIMAL WEIGHTS OF SSNR AND SSIM.

	Weights
SSNR	1.5552
SSIM	22.6861

B. The Hybrid Metric

The global coefficients ($\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*$) and optimal W_{Global} obtained are shown in Table V and Table VI, respectively. Fig. 11 demonstrates that the correlation coefficients of coarse materials (*Coarse Board* and *Jeans*) are higher than that of fine materials (*Glass* and *Leather*) for either SSNR or MSSIM. It suggests that SSNR and MSSIM perform slightly better for assessing time-varying distortion of coarse materials. As to stiffness, MSSIM is more correlated with soft materials, while SSNR does not show apparent preferences. As for the hybrid metric, the correlation coefficients of all types of materials are quite similar and do not show any obvious tendencies. This is because it takes advantages of both SSNR and MSSIM. Comparing the hybrid metric with SSNR and MSSIM, we can see an improvement of about 0.08 in terms of the average correlation coefficients. In other words, the hybrid metric achieves better performance than SSNR and MSSIM in the time-varying framework. Moreover, it is expected that the performance of the hybrid metric will further improve as more objective metrics for vibrotactile data are introduced in the future. As for the significance test, the degree of freedom is 10 in this case as there are 12 samples for each type of material. Since the average correlation coefficient of the hybrid metric, i.e. 0.8947, is larger than the critical value at significance level of 0.01, i.e. 0.7079, shown in the table of critical values. We can thus say that the correlation coefficient of the hybrid metric is significant within a 99% confidence interval.

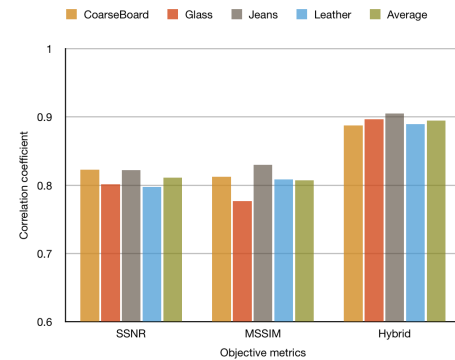


Fig. 11. Pearson correlation coefficients in the time-varying testing framework. The correlation coefficients of the three objective metrics with respect to different materials are depicted.

VII. CONCLUSIONS AND FUTURE WORK

Firstly, we empirically demonstrate that the SNR and SSIM are applicable to vibrotactile quality assessment. As objective metrics can only be validated through subjective experiments, we propose a subjective protocol, referred to as MUSHRT, adapted from Rec. ITU-R BS.1534-3 in the first place. Based on this protocol, subjective experiments are conducted and the results demonstrate that SNR and SSIM are highly correlated with human vibrotactile perception. In addition, SSIM outperforms SNR, which leads to a possible direction of future study, i.e. designing vibrotactile SNR that fits human subjective scores.

After validating SNR and SSIM, we consider a realistic scenario with the quality of vibrotactile data varying with time. In the light of PEAQ, we design a time-varying metric by combining SNR and SSIM and evaluate its performance through subjective experiments. The results show that an increase of 8% in terms of correlation coefficient is achieved.

In summary, the vibrotactile quality assessment is in the initial stage. It is expected that more objective metrics for vibrotactile data will be developed in the future. We can hence utilise more MOVs as inputs of the hybrid metric design, which is likely to further improve the performance. Another direction of future work is to develop quality assessment for more realistic scenarios where traditional multimedia is combined with vibrotactile data. Furthermore, the quality evaluation in this article considers vibrotactile data obtained through sliding. It is worth investigating other types of vibrotactile data, such as vibrations obtained through tapping, in the future.

ACKNOWLEDGMENT

The authors would like to thank Matti Strese in Technical University of Munich for his kind help of setting up the LVCA.

REFERENCES

- [1] G. Ghinea and O. Ademoye, "A user perspective of olfaction-enhanced mulsemmedia," in *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*. ACM, 2010, pp. 277–280.
- [2] Z. Yuan, G. Ghinea, and G. Muntean, "Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 104–117, Jan 2015.
- [3] M. Walzl, C. Timmerer, and H. Hellwagner, "Improving the quality of multimedia experience through sensory effects," in *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2010, pp. 124–129.
- [4] J. Monks, A. Olaru, I. Tal, and G. M. Muntean, "Quality of experience assessment of 3d video synchronised with multisensorial media components," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 2017, pp. 1–6.
- [5] N. Murray, G. M. Muntean, Y. Qiao, S. Brennan, and B. Lee, "Modeling user quality of experience of olfaction-enhanced multimedia," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 539–551, 2018.
- [6] N. Murray, B. Lee, Y. Qiao, and G. M. Muntean, "The impact of scent type on olfaction-enhanced multimedia quality of experience," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 9, pp. 2503–2515, 2017.
- [7] D. A. Lawrence, "Stability and transparency in bilateral teleoperation," *IEEE transactions on robotics and automation*, vol. 9, no. 5, pp. 624–637, 1993.
- [8] S. Okamoto, M. Konyo, S. Saga, and S. Tadokoro, "Detectability and perceptual consequences of delayed feedback in a vibrotactile texture display," *IEEE Transactions on Haptics*, vol. 2, no. 2, pp. 73–84, April 2009.
- [9] P. Hinterseer, S. Hirche, S. Chaudhuri, E. Steinbach, and M. Buss, "Perception-based data reduction and transmission of haptic data in telepresence and teleaction systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 588–597, 2008.
- [10] N. Sakr, N. D. Georganas, and J. Zhao, "A perceptual quality metric for haptic signals," in *2007 IEEE International Workshop on Haptic, Audio and Visual Environments and Games*, Oct 2007, pp. 27–32.
- [11] R. Chaudhari, E. Steinbach, and S. Hirche, "Towards an objective quality evaluation framework for haptic data reduction," in *2011 IEEE World Haptics Conference*. IEEE, 2011, pp. 539–544.
- [12] R. Hassen and E. Steinbach, "Hssim: An objective haptic quality assessment measure for force-feedback signals," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018, pp. 1–6.
- [13] J. Maiero, E. Kruijff, A. Hinkenjann, and G. Ghinea, "Forcetab: Visuo-haptic interaction with a force-sensitive actuated tablet," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 169–174.
- [14] ITU-R, Recommendation BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," *International Telecommunications Union, Geneva Switzerland*, 1994.
- [15] Z. Yuan, T. Bi, G. Muntean, and G. Ghinea, "Perceived synchronization of mulsemmedia services," *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp. 957–966, July 2015.
- [16] S. Okamoto and Y. Yamada, "Perceptual properties of vibrotactile material texture: Effects of amplitude changes and stimuli beneath detection thresholds," in *IEEE/SICE International Symposium on System Integration (SII)*, 2010, pp. 384–389.
- [17] R. Chaudhari, B. Çizmeçi, K. Kuchenbecker, S. Choi, and E. Steinbach, "Low bitrate source-filter model based compression of vibrotactile texture signals in haptic teleoperation," in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 409–418.
- [18] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Human Vision and Electronic Imaging X*, vol. 5666. International Society for Optics and Photonics, 2005, pp. 149–160.
- [19] S. Okamoto and Y. Yamada, "An objective index that substitutes for subjective quality of vibrotactile material-like textures," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 3060–3067.
- [20] X. Liu, M. Dohler, T. Mahmoodi, and H. Liu, "Challenges and opportunities for designing tactile codecs from audio codecs," in *2017 European Conference on Networks and Communications (EuCNC)*, June 2017, pp. 1–5.
- [21] R. Chaudhari, C. Schuwerk, M. Danaei, and E. Steinbach, "Perceptual and bitrate-scalable coding of haptic surface texture signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 462–473, 2015.
- [22] E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, and A. Hudspeth, *Principles of Neural Science, Fifth Edition*. McGraw-Hill Education, 2012.
- [23] J. Bell, S. Bolanowski, and M. H. Holmes, "The structure and function of pacinian corpuscles: a review," *Progress in neurobiology*, vol. 42, no. 1, pp. 79–128, 1994.
- [24] R. Klatzky and S. Lederman, "Perceiving object properties through a rigid link," *Haptic Rendering: Algorithms and Applications*, vol. 1, pp. 7–19, 2008.
- [25] T. Yoshioka and J. Zhou, "Factors involved in tactile texture perception through probes," *Advanced Robotics*, vol. 23, no. 6, pp. 747–766, 2009.
- [26] A. M. Okamura, M. R. Cutkosky, and J. T. Dennerlein, "Reality-based models for vibration feedback in virtual environments," *IEEE/ASME Transactions on Mechatronics*, vol. 6, no. 3, pp. 245–252, 2001.
- [27] K. J. Kuchenbecker, J. Fiene, and G. Niemeyer, "Improving contact realism through event-based haptic feedback," *IEEE transactions on visualization and computer graphics*, vol. 12, no. 2, pp. 219–230, 2006.
- [28] W. McMahan, J. M. Romano, A. M. A. Rahuman, and K. J. Kuchenbecker, "High frequency acceleration feedback significantly increases the realism of haptically rendered textured surfaces," in *Haptics Symposium, 2010*. IEEE, 2010, pp. 141–148.
- [29] ITU-T, Recommendation P.800, "Methods for objective and subjective assessment of quality," *International Telecommunication Union, Geneva, Switzerland*, 1996.
- [30] ITU-R, Recommendation BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems," *International Telecommunication Union, Geneva, Switzerland*, 2014.

- [31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [32] D. M. Chandler and S. S. Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE transactions on image processing*, vol. 16, no. 9, pp. 2284–2298, 2007.
- [33] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2117–2128, 2005.
- [34] W. Yang, *Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measure Based on Audible Distortion and Cognitive Model*. Temple University, 1999.
- [35] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, vol. 2, May 2001, pp. 749–752 vol.2.
- [36] ITU-R, Recommendation BS.1387-1, "Method for objective measurements of perceived audio quality," *International Telecommunication Union, Geneva, Switzerland*, 2001.
- [37] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [38] S. Kandadaï, J. Hardin, and C. D. Creusere, "Audio quality assessment using the mean structural similarity measure," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 221–224.
- [39] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [40] P. Mermelstein, "Evaluation of a segmental snr measure as an indicator of the quality of adpcm coded speech," *The Journal of the Acoustical Society of America*, vol. 66, no. 6, pp. 1664–1667, 1979.
- [41] G. T. Fechner, D. H. Howes, and E. G. Boring, *Elements of psychophysics*. Holt, Rinehart and Winston New York, 1966, vol. 1.
- [42] J. Kirsch, A. Noll, M. Strese, Q. Liu, and E. Steinbach, "A low-cost acquisition, display, and evaluation setup for tactile codec development," in *2018 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, 2018, pp. 1–6.
- [43] N. Landin, J. M. Romano, W. McMahan, and K. J. Kuchenbecker, "Dimensional reduction of high-frequency accelerations for haptic rendering," in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 2010, pp. 79–86.
- [44] The H2W Technologies. [Online]. Available: <https://www.h2wtech.com>
- [45] Velleman. [Online]. Available: <https://www.velleman.eu>
- [46] The Bostik Adhesive. [Online]. Available: <https://www.bostik.com>
- [47] G. Waters, "Sound quality assessment material-recordings for subjective tests: User's handbook for the ebu-sqam compact disk," *European Broadcasting Union (EBU), Tech. Rep.*, pp. 1–13, 1988.
- [48] S. Zielinski, P. Hardisty, C. Hummersone, and F. Rumsey, "Potential biases in mushra listening tests," in *Audio Engineering Society Convention 123*. Audio Engineering Society, 2007.
- [49] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality: an explanation of $1/f$ noise, 1987," *Phys. Rev. Lett.*, vol. 59, p. 381.



Mischa Dohler Mischa Dohler is full Professor in Wireless Communications at King's College London, driving cross-disciplinary research and innovation in technology, sciences and arts. He is a Fellow of the IEEE, the Royal Academy of Engineering, the Royal Society of Arts (RSA), the Institution of Engineering and Technology (IET); and a Distinguished Member of Harvard Square Leaders Excellence. He is a serial entrepreneur; composer&pianist with 5 albums on Spotify/iTunes; and fluent in 6 languages.

He acts as policy advisor on issues related to digital, skills and education. He has had ample coverage by national and international press and media. He is a frequent keynote, panel and tutorial speaker, and has received numerous awards. He has pioneered several research fields, contributed to numerous wireless broadband, IoT/M2M and cyber security standards, holds a dozen patents, organized and chaired numerous conferences, was the Editor-in-Chief of two journals, has more than 200 highly-cited publications, and authored several books. He was the Director of the Centre for Telecommunications Research at King's from 2014-2018. He is the Cofounder of the Smart Cities pioneering company Worldsensing, where he was the CTO from 2008-2014. He also worked as a Senior Researcher at Orange/France Telecom from 2005-2008.



Yansha Deng Yansha Deng (S'13-M'18) received the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2015. From 2015 to 2017, she was a Post-Doctoral Research Fellow with King's College London, U.K., where she is currently a Lecturer (Assistant Professor) with the Department of Informatics. Her research interests include molecular communication, machine learning, and 5G wireless networks. She was a recipient of the Best Paper Awards from ICC 2016 and Globecom 2017 as the first author. She is

currently an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE COMMUNICATION LETTERS. She also received an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2016 and 2017. She has also served as a TPC Member for many IEEE conferences, such as IEEE GLOBECOM and ICC.



Xun Liu Xun Liu is a PhD student in the Department of Informatics at King's College London in United Kingdom. He obtained B.Eng with first honour from University of Birmingham in 2013 and MSc with distinction from Imperial College London in 2014. He is doing research on tactile internet and focuses on designing codecs and quality evaluation metrics for tactile signals. He is a student IEEE member. He is also a voting member of IEEE P1918.1 Tactile Internet Working Group and contributes to development and standardisation of

Tactile Internet.